AD_____

GRANT NUMBER DAMD17-94-J-4262

TITLE: Childbearing and Survival After Breast Cancer

PRINCIPAL INVESTIGATOR: Beth A. Mueller, Ph.D.

CONTRACTING ORGANIZATION: Fred Hutchinson Cancer Research Center
Seattle, Washington 98104-2092

REPORT DATE: September 1996

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1996	Annual (30 Aug 95 - 29 Aug 96)

4. TITLE AND SUBTITLE
Childbearing and Survival After Breast Cancer

5. FUNDING NUMBERS
DAMD17-94-J-4262

6. AUTHOR(S)
Beth A. Mueller, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Fred Hutchinson Cancer Research Center
Seattle, Washington 98104-2092

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 2170-25012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

The incidence of breast cancer among women less than 45 years of age is increasing, and young women appear to have relatively poor survival. The prognosis may be even worse for women who are pregnant at diagnosis; however, the effect on survival of a pregnancy conceived after breast cancer is unknown. As the increase of breast cancer incidence among younger women coincides with a trend towards delayed childbearing, information regarding the association of subsequent pregnancy and survival is needed so that women with breast cancer and their physicians can make informed choices concerning family planning. We propose to determine whether women with breast cancer who subsequently bear children are less likely to survive than a matched comparison group of women with breast cancer who do not bear a child.

We are in the process of linking cancer registry data in three US populations with birth certificate records. Women will be followed up through 1995 to determine survival.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	21
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

BAM X For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Beth Mueller 9/25/96

PI - Signature Date

Table of Contents

I. INTRODUCTION

Our study scope and proposal have not changed over the previous year.

A. Nature of the problem

The incidence of breast cancer among women less than 45 is increasing, and young women appear to have relatively poor survival. The prognosis may be even worse for women who are pregnant at diagnosis however, the effect on survival is unknown . As the increase in breast cancer incidence among younger women coincides with a trend towards delayed childbearing, information regarding the association of subsequent pregnancy and survival is needed so that women with breast cancer and their physicians can make informed choices concerning family planning.

At present there is no general consensus among physicians providing care to young women with breast cancer about how to advise them regarding their future reproduction (14), and the lay press clearly demonstrates that this is an issue of concern to those women affected. Current clinical recommendations concerning a waiting period of 2-3 years (the peak period of recurrence) after the conclusion of breast cancer treatment before attempting pregnancy are based on psychosocial and moral issues rather than scientific studies linking pregnancy with poorer survival(3,14). Women who survive their initial breast cancer treatment are justifiably confused concerning their future.

B. Previous work

The effect on survival of a pregnancy conceived after breast cancer treatment has been little studied and the available results have only incompletely accounted for confounding and selection bias. There are several reasons why few studies have evaluated the effect of subsequent childbearing on survival among women with breast cancer. First, although breast cancer is one of the most common of female cancers, young women in their childbearing years account for a relatively small percent of all primary cases (1). Second, only a minority of them will go on to conceive and carry a pregnancy to term afterwards(2). Thus it is difficult to identify and study a large enough number of women to obtain valid results.

Survival may be relatively worse for women who are pregnant at the time of their primary diagnosis of breast cancer, although it is difficult to generalize from previous studies because of varying definitions and populations used. Studies of pregnancy-associated breast cancer may not distinguish between women diagnosed during pregnancy soon after pregnancy, during lactation, lactation, or even subsequent to treatment. Most studies have included women who have undergone mastectomy only, and the time periods covered have encompassed wide variation and changes in clinical treatment. Details of treatment, which are likely to have an impact on survival, are rarely considered in the analysis (3). Early reports suggested that breast cancer diagnosed

during pregnancy was associated with poor survival (3,4), however, this observation may be confined to node-positive breast cancer diagnosed during gestation. Recent studies that have controlled for the effects of age and stage at diagnosis however, have conflicting results. Peters (5), observed no difference in stage between women diagnosed during pregnancy or lactation and a control series matched on age and stage of disease. These results have been supported by at least one other study (6). which also suggested that young age (less than 40 years), rather than pregnancy itself may be the negative prognostic factor responsible for previous studies suggesting poor survival in pregnant women, as young women in their study had a relatively higher proportion of aggressive estrogen receptor-negative tumors than older women. Petrek et al. (7) followed up 56 pregnant women treated surgically at a single facility, and a similar group of 166 non-pregnant patients, and reported similar 5 year survival (82%) among node-negative women who were pregnant and nonpregnant. Node-positive women who were pregnant, however, had slightly worse 5 year survival (47%) than node-positive nonpregnant women (59%). Two other recent studies that controlled for age and disease stage, however, suggest a significantly worse survival among women diagnosed during pregnancy (8-9). There have been several reasonable biological rationales suggested to support an association of pregnancy-associated breast cancer with poorer survival, including variations in hormone levels and immunologic changes such as a decrease in cell-mediated immunity during pregnancy (3,12). Uncovering the true association is further complicated, however, as the elapsed time from detection to diagnosis is frequently longer in pregnant women than non-pregnant women, which may account, to some extent, for a relatively poorer survival (3,12).

An early study (1937) reported that women with pregnancies after surgery had better survival than those whose breast cancer was diagnosed during pregnancy or lactation(4). Several subsequent studies that have compared women with pregnancies after breast cancer diagnosis to comparison groups without pregnancies suggest no difference in survival. Peters(5) reported 72% vs. 50% 5 year survival in patients with and without subsequent pregnancy, even after controlling for age and stage at diagnosis. When only patients who became pregnant within 6 months of surgery were included, however the proportion of women with subsequent pregnancies surviving at 5 years dropped to 54% indicating the need to control for bias due to the decreased likelihood of pregnancy among women with more rapid disease progression. One study matched 32 patients with pregnancies to two controls (on the basis of clinical stage, node involvement, age, and initial survival after surgery) and concluded that pregnancy did not adversely affect survival (10). However, they only followed patients for 5 years after diagnosis and did not include patients with other treatment modalities. Another study followed up radical mastectomy cases without chemotherapy from 1930-1975 at one facility (47 with pregnancies/960 without pregnancies) and reported no differences in 5 year survival. Among node-negative cases, 5 year survival was 77% for women with pregnancies vs. 70% for those without; among node-positive cases, it was 56% vs. 53% (11). The authors conclude that pregnancy after treatment of breast cancer has no effect upon prognosis, however, their study included only women with radical mastectomies

treated at a single facility and could not address pregnancy and survival after other treatments.

A recent Swedish study (1995) compared women diagnosed with breast cancer without pregnancies to women with breast cancers and either prior (N=173) or subsequent pregnancies (N=50) within 5 years of diagnosis. This study compared all incidences of primary operable breast cancer diagnosed in the area of Stockholm between 1971 and 1988. The authors studied tumor size, ER status, age, and found results that supported Petrek's earlier conclusion that pregnancy had no adverse effect on the prognosis of breast cancer (13). They also concluded that a subsequent pregnancy may be related to a decreased risk of distant dissemination. (relative hazard = 0.48, p=.14). An examination of this with a greater number of subjects, longer follow-up, and with an examination of treatment modalities such as we are conducting, needs to occur.

C. Purpose of present work

The purpose of the present study is to obtain data from a large enough sample size, and with a diverse enough population, in order to evaluate the relative survival of women with and without births among young women with breast cancer. By obtaining data from three population-based cancer registries (Seattle, Los Angeles, and Detroit) we can identify all women <45 years, with breast cancer. Similar clinical information is obtained from all registries, thus it will be possible to evaluate the effects of stage, node status, and other important factors on survival. Using birth certificate data from all three states (Washington, California, and Michigan) we will ascertain those with and without births in order to form the exposed and comparison groups. Additional data obtained from previous case-control studies conducted in Los Angeles and Seattle will allow subanalyses with information related to family history, body mass, estrogen/oral contraceptive exposures and other data that may potentially influence the relationship of interest. In addition, the use of data from three such ethnically diverse areas will allow us to further examine this relationship in a way that it has not been done before.

D. Methods of approach

Our study makes use of three population based cancer registries and birth certificate records in their respective states in order to determine the subsequent childbearing experience of young women with primary breast cancer. Women <45 years of age at diagnosis of breast cancer have been identified in the Seattle area, Detroit, and Los Angeles cancer registries. Data tapes containing birth certificate information from 1980-1994 from all three states will be merged with cancer registry information to identify women with subsequent births. As all registries routinely link to death certificate information, it will be possible to screen subjects for survival.

After linkage of registry data to birth certificates and identification of women with births, a comparison group of women without births will be identified. They will be drawn from among women <45 years of age with breast cancer who were found not to have had births and will be linked on the basis of relevant characteristics (age, race, stage of disease at diagnosis, year of diagnosis) and will be selected to consist of only those who have survived up until the reference date (date of pregnancy for women with births) of their respective match. This latter step will be done to ensure the relative health of both the exposed (those with births) and comparison groups. Both groups will be similarly followed up for survival and disease recurrence using routine methods at the cancer registries. The survival of women with births will be compared to that of the comparison group us Cox Proportional Hazards Regression analysis, a technique that will allow us to control more completely for other factors related to survival, including nodal status, treatment, and estrogen receptor status, in evaluating relative survival after childbirth.

II. BODY

describe experimental methods and results obtained relative to goals of research

A. Experimental methods used

To achieve our objective, we are using a retrospective cohort design, comparing the survival of young women who give birth after a diagnosis of breast cancer with that of a matched comparison group of breast cancer patients who do not give birth. Emphasis will be placed on determining the disease status of women at the time the pregnancy occurred (one calendar year prior to the birth, or the "reference date") in an attempt to ensure that the health status of the women with and without births are similar. Young women less than 45 years of age with breast cancer diagnosed during 1980-1993 from three population-based cancer registries (Seattle, Los Angeles, Detroit) have been identified. This cohort file is being linked with birth certificate records for 1980-1993 in each state to identify women who have a live birth after their diagnosis of primary breast cancer. Subjects eligible for inclusion in the comparison group will be selected from among women without births at each site who are known to be alive at the reference date, and will be matched on age, stage at diagnosis, and year of diagnosis. All women will be followed up using routine methods to determine their survival after reference date through 1995 (follow-up ranging from 5 to 15 years, with an expected mean follow-up of approximately 7.7 years).

After creation of the data sets containing registry and birth certificate information from each site, data will be consolidated into one research data set in Seattle, where data analysis will occur. Cox proportional hazards regression will be used to evaluate the relative survival after reference date of women with and without subsequent births. This method allows for adjustment in the analysis of potentially confounding variables obtained from registry data (tumor size, nodal status, estrogen receptor status).

B. Results obtained relative to our stated goals

Per our revised Statement of Work, our second year of operation has consisted of activities related to data linkage and the selection of the matching control cohort.

Communications between the sites occurred frequently and the investigators from all Sites (Drs. Mueller, Deapen, and Simon and the Data Coordinator, Janet Kelly) met in the Spring in Portland, Oregon to discuss study issues and to apprise each other of progress. A copy of the points of discussion at this meeting is in appendix A. The meeting included discussions of the linkage procedure at each site, the results of running the

control matching program and issues relating to the relaxation of criteria for matching controls to cases on age, race, year of diagnosis, stage at diagnosis.

Also discussed was the need to explore the possibility of obtaining the fetal death records from the respective sites departments of health. The fetal death records would help us identify women who became pregnant after being diagnosed with breast cancer but were unable to carry the pregnancy to term. Without these records, we may miss linking a portion of women who were pregnant but did not give birth and therefore bias our study toward only those women able to deliver a live birth. Fetal Death records are those deaths that occurred >20 weeks and are not induced abortions. The fetal death records are kept in the same format as the birth certificate records. Seattle has received all necessary approvals from the Washington State Department of Health and the Fred Hutchinson Cancer Research Center's IRB to use the fetal death records. There is no extra cost for obtaining the fetal death records at the Seattle site and using these records does not change the scope of the study.

Seattle has conducted the fetal death linkage with the CSS records to ascertain how many links are gained by this method. For the years of data 1980-1994 we found only one link between the cancer registry data and the fetal death records. Given this low number of linkages found, it was decided that it would not be worthwhile to proceed in obtaining the fetal death records at the other study sites.

Activities which occurred during the last year are described below in more detail:

1. Identification of young women with breast cancer at all sites - At all three sites, women <age 45 diagnosed with breast cancer during the years 1980-1993 have been identified. In Seattle, 3,925 women who fit the preliminary inclusion criteria have been identified and 82 have been linked with the birth certificate tapes. In Detroit, 4,496 women have been identified who meet study criteria. In L.A. 6,962 women met the criteria.

2. Procurement of birth certificate data tapes - Appropriate permission has been obtained by the investigator, and the co-investigators at all sites, from all relevant state Departments of Health and Human Subjects Protection Committees for use or purchase of birth certificate data files for 1980-94. All the necessary tapes have been located and accessed by the investigators at each site.

3. Development of linkage protocols-
In the previous study year, the necessary birth records and the cancer registry data were obtained. Protocols for each site were developed to ensure compliance with each sites' department of health. Once that was accomplished, work on the actual linkages occurred. Computer programs which generated a list of possible links between the cancer registry data and the birth records were written and run at each site. Due to the each sites' differing department of health requirements, the programs were slightly

different although they used the same general programming code to ensure the same linking criteria.

Once the list of possible links was generated by the computer, a member of the investigative team hand checked the list to verify the linkages and weed out any links that were not a probable link. The criteria to hand check the links are attached in appendix B. Although some refinement may occur in these numbers, to date it appears that the number of links (women with a birth ten or more months after breast cancer diagnosis) at the Seattle site is 3.6% of the total possible cohort. In Detroit the percentage of links is 5.0%. In L.A. the percentage link is 4.0%. These numbers are fairly close to the number of links expected.

The L.A. site independently conducted a comparison of results from our SAS matching program to those obtained from a matching program based on probability methods in order to validate our procedure. Preliminary comparison indicates that a similar set of matches is obtained with both methods.

We now have an exposed cohort in each site with which to work. Below is a table describing the cohort at the Seattle site, which is spearheading the analysis.

Characteristics of Young Women Diagnosed with Breast Cancer Who Linked to Subsequent Years of Birth Certificate Data.

Seattle Site	***Mean+SD***
Mean Age at Diagnosis±SD	*31.4±4.0*
Mean Age of Mother at time of Birth ±SD	*34.7±4.0*
Mean number of months between diagnosis and birth± SD	*40±27.2 or 3.4 years*

	N	***Percentage***
Race		
- White	*75*	*91.5*
- African- American	*4*	*4.9*
- Filipino	*1*	*1.2*
- Vietnamese	*1*	*1.2*
- Unknown	*1*	*1.2*
Year of Dx		
1980-84	*26*	*31.7*
1985-89	*37*	*45.1*
1990- 93	*19*	*23.2*
Stage at DX		
Local	*55*	*67.1*
Regional	*26*	*31.7*
Unknown	*1*	*1.2*
Total	*82*	

4. Identify reference dates for women in exposed cohort

A preliminary program to match the exposed cohort to the cases has been written and run at each site. At the meeting in Portland, some concerns arose about the matching program and the need for the program to be modified was made clear. However, the program was run to match the cases to the controls as a way to find how appropriate the matching criteria (age, stage at diagnosis, year of diagnosis, and race) were. The computer program that matches the control cohort with the exposed cohort is being refined and modified to match the exposed cohort based on the reference dates.

5. Select 5 comparison candidates from unexposed women based on reference date and matching variables.

The preliminary matching program was run at each site. Based on the results of this program, it will be difficult to find at least five matching controls for the younger cases, especially in the advanced stages of breast cancer. Although it is possible to match the older cases with at least five controls, the younger cases many times had none. To alleviate this problem a protocol of relaxation of the criteria was developed in part through the discussions in Portland and through research. It was decided at this meeting that Stage at Diagnosis and Race would not be relaxed. Age at diagnosis may only be relaxed -2 years, and year of diagnosis may be categorized into 1980-1984, 1985-1989, 1990 and greater.

Based on this relaxation of criteria, when the preliminary control matching program was run, at least one control matched one case. However, due to the makeup of the unexposed cohort, the number of controls matching the cases varied greatly.

6. Refine comparison cohort based information from survival follow-up.
Refinement of the comparison cohort continues and will be modified throughout the coming year.

III. CONCLUSIONS

A. Implications of work completed to date and description of changes made to procedures

Work is being completed on schedule. Next year will be devoted to selection of comparison group and analyses will be underway within six months. Next year will also include work on the follow up procedures to find up to date vital status of the women. The sources of the follow-up information will likely include searching death certificates, follow up information from current studies, the use of telephone and reverse directories, and possibly, depending on IRB regulations at each site, contact with the department of motor vehicles.

There will be another meeting in the Spring of 1997 and continued close communication between investigators via email and telephone.

No major changes have been made to the protocol. We did evaluate the possibility of examining the fetal deaths as an outcome, however. Although it is feasible, this was determined to be a low yield activity based on preliminary linkage of Seattle data.

Bibliography

1. Miller BA, Feurer EJ, Hankey BF. Recent incidence trends for breast cancer in women and the relevance of early detection: An update. CA Cancer J Clin 1993; 3:27-41.
2. Donegan WL. Mammary carcinoma and pregnancy. in: Donegan WL, Spratt JS, Eds. Cancer of the Breast. Philadelphia: WB Saunders, 1985: 679-3.
3. vanderVange N, vanDongen JA. Breast cancer and pregnancy. Eur J Surg Oncol 1991; 17(1):1-8.
4. Harrington SW. Carcinoma of the breast. Ann Surg 1937; 106:690-700. 688.
5. Peters MV. The effect of pregnancy in breast cancer and pregnancy. in: Forrest AP, Kunkler PB, Eds. Prognostic Factors in Breast Cancer. 14.Baltimore: Williams & Williams, 1968:65-80.
6. Nugent P, O'Connell TX. Breast cancer in pregnancy. Arch Surg 1985; 120(11): 1221-4.
7. Petrek JA, Dukoff R, Rogatko A. Prognosis of pregnancy-associated breast cancer. Cancer 1991; 67(4): 869-872.
8. Mailer WP. Carcinoma of the breast in pregnancy and lactation. Breast 1979; 4:12-15.
9. Tretli S, Kvalheim G, Thoresen S, Host H. Survival of breast cancer patients diagnosed during pregnancy or lactation. Br. J Cancer 1988; 58(3): 382-384.
10. Cooper DR, Butterfield J. Pregnancy subsequent to mastectomy for cancer of the breast. Ann Surg 1970;171:429-433.
11. Ariel IM, Kempner R. The prognosis of patients who become pregnant after mastectomy for breast cancer. Int Surg 1989; 74:185-187.
12. Gallenberg MM, Loprinzi CL. Breast cancer and pregnancy. Semin Oncol 1989; 16(5):369-376.
13. von Schoultz E, Johnasson H , Wilking N, Rutqvist LE. Influence of prior and subsequent pregnancy on breast cancer prognosis. J Clin Oncol 1995; 13:430-434.
14. Saunders CM, Baum M. Breast cancer and pregnancy: a review. J Royal Soc Med 1993; 86(3): 162-165.

Appendix A

Meeting Notes for the Childbearing and Survival Study:
12:00 to 3:00 June 4th, 1996
Shilo Inn, Portland Oregon
Prepared by Janet Kelly

Issues Discussed and Actions Needed (not in chronological order).

1). Cohort will include those women who gave birth >=10 months after dx. Women in the <10 months or prior to dx, may be identified but will not be included in the principle analysis dataset.

2). Fetal Death Records.
To obtain matches of women with breast cancer who became pregnant but did not carry the pregnancy to term, Fetal Death Records would need to be matched with the cancer registries.

Since obtaining the records requires additional approval from the states and money to buy the tapes, Seattle (which already has access to these tapes) will do a preliminary run to see how many additional links are gained. If it is deemed worthwhile to spend the money and the time to get the tapes for LA and Michigan based on the number of additional links gained in Seattle, we will proceed with obtaining the Fetal Death Records for the other sites. At present, Dennis and Michael will explore the costs and feasibility of obtaining the tapes in their states.

3). CSS Registry and Birth Certificate Linking Issues
a). The need for a documented description of the Human Decision Making Process for the links was discussed. Specifically, after the computer linking program is run, what is necessary for the human, who is hand checking the computer links, to make a decision about a definite link, possible link, and no link. Janet will write a more thorough guideline for this.

b). Also discussed was the need to look at CSS record abstracts for more information about unknown race or stage for the linking purposes. Once linked, if race is unknown in the CSS record but known on the birth certificate, we may replace that unknown CSS race with the birth certificate race in the analysis data. This may allow us to achieve a higher percentage of matches. Every change in data must be documented. Janet will develop a standardized form to accommodate these changes. If stage of disease is unknown and cannot be obtained, the record will be excluded from the final analysis.

Include all possible matches for now.

4). Control Matching Issues.
The main issue discussed about matching controls with cases by age, year of dx, stage, race, (and alive w/o recurrence at the time of birth) was how to gain the most controls matched per case. Even after all criteria are relaxed (exacts, age -2 years, dx year$\pm$2 race into white and non white) some cases had no matching controls, particularly in younger women (less than 30 years old).

Issues discussed regarding the relaxation of criteria were as follows:
a). If race is relaxed into white vs. non white, is there an issue of survivability differences between non white races, i.e. African American vs. Asians? Relaxing race in this way was not thought to be acceptable and exact race matching will be retained.

b). If year of diagnosis is relaxed into year categories (80-84, 85- 89, 90+), is survival different by years of diagnosis? The SEER Report: Stat Review, Survival of Breast Cancer Tables, will be examined to see if these ranges are acceptable.

c). Because of the difficulty matching younger women, we discussed relaxing age into <=35 $\pm$2 years and >35- 2 years. This however was deemed inappropriate due to the disease's course variations on age and the confusing nature of the criteria relaxation. The possibility of relaxing only age -2 years was discussed and will be reevaluated after we see the percent matches when dx year is relaxed.

d). Due to the pool of possible controls, the program to match the controls will select a variable number of controls per case. It was decided to at first, take every possible match. However, we may refine this later.

e). The program to match the controls will be refined to handle the situation if 2 cases have identical criteria (cmatch number) to ensure that the controls are chosen randomly. All controls should be left in the pool to sample with replacement at least initially to make possible random selection from all potential control candidates. The program will be refined to handle these issues.

f). Case/Controls will be matched at all three sites before being merged in Seattle.

g). To describe and document how the criteria were relaxed, Janet will develop a case/control match variable of how well control matches link (codes will show if and what criteria was relaxed).

5). Beth has given a bibliography of published articles about linkages to Detroit. These articles will be needed for work on a linkage methodology paper.

6). Other Information needing to be obtained.
From medical information about the birth: Seattle will develop a uniform data list of variables needed for all three sites.

7). LA reported no routinely obtained follow up information available on vital status (with the exception of death and any standard SEER follow up data) until 1992. However, Dennis mentioned the existence of his Tracing Lab, developed for another project, which can conduct more extensive follow up.

Appendix B

Selection Criteria for Childbearing Matching After Breast Cancer Linkage. Revised 6/3/96 and 9/16/96.

These guidelines are to be used after computer linking program is run as a means of verification of results.

Printouts are checked for accuracy and verified by hand. The person checking (from here on, called the link checker) these printouts must make a final decision about the quality of a potential link. The following guidelines are designed not only to aid the link checker in their decision making process, but serve as a standard coding scheme for all study sites.

Definition of a "hit"

Some of the links between the cancer registry data and the birth certificate data will be obvious and considered to be a link beyond a reasonable doubt.

1). To qualify as a "hit" (a link of beyond a reasonable doubt) the potential matches of the cancer registry record and the birth certificate record need to pass certain criteria. For our purposes, we will define a link as a record in which the cancer registry data last name, first name, maiden name (or middle name), and date of birth match the birth certificate record last name, first name, maiden name (or middle name) and date of birth.

2). For years that dates of birth are available and the name is an exact match, it allows us the opportunity to check for data entry errors. Many times six or seven out of the eight digits of the dates of birth on both records match. Where this is the case, it may be considered a probable link provided that the address can be verified to indicate that it is the same person.

3). When the of birth is not available, but age is, age may be substituted for date of birth provided that the addresses on both the cancer registry record and the birth certificate record can be checked and verified.

Assumptions made about the data and the quality of the link

While matching on name, date of birth, age, and address, there are some assumptions that need to be made by the link checker about the data in both the cancer registry and the birth certificate data in order to make informed decision about a potential link. We need to assume that all of the information contained in that record is accurately collected and entered correctly (without data entry errors) into the respective databases. While isolated cases of data entry errors may occur, the overall assumption is that the data is correct. We also need to assume that the records contain information gathered by conventional North American reporting methods. These two basic assumptions are needed in order to work effectively with data records from differing sources.

The need to have these assumptions defined as part of the record linkage process becomes clearer as certain situations arise while the link checker works with the data.

An exact name match of a common name (for example Mary Johnson) and a date of birth that is a close match (perhaps five of eight digits match) would not be considered a link due to the assumption that the data in both records is accurate and entered correctly. If there was no assumption, this situation may be considered a possible link, although not a definate link beyond a reasonable doubt. Both sources of the dates of birth assume their information is the actual date of birth. In this situation, assuming that all information is accurate, the bottom line is that we have a name match and a different date of birth (although some digits match - not enough of them do) which, strictly defined is not a link.

Another common situation involves of immigrants from other cultures living in North America that may not share the same naming and dating conventions that North American institutions use. This immigrant contingent while using North American institutions (hospitals) may not understand our conventions or know that there is a difference in naming and dating conventions between cultures. Specifically, this comes in to play with Asian cultures that use the date of conception for the date on which to base the age instead of the actual date of birth as it is in North America. Since our data is dependant on time, this difference of nine months would drastically alter the data. However, there is no way of knowing that might be the situation just by looking at the information in the linkage records. Therefore, we have to "assume" that every record in the cancer registry data and the birth certificate data is using the standard North American dating conventions due to the fact they are coming from North American Institutions.. The same could be said of the naming conventions. Women who do not speak English may not understand the concept of last name, first name when filling out North American hospital forms. They may enter the last name in the first name column. But, again, we need to assume the last name is really the last name in all records.

There are additional relatively minor and obvious assumptions which are also necessary to define. This is where all other criteria and assumptions do not go far enough to define all possible scenarios. These assumptions don't necessarily indicate a non match, but they do provide an aid for making a decision. If any of the following assumptions are not followed, the link variable needs to be coded as a possible link, as opposed to a definate link. We are defining the minor assumptions as: a situation in which a record of a potential link may not hold all of the information that is needed by the link checker to make an informed decision beyond a reasonable doubt. A possible link is a link that includes a common sence assumption for lack of other information.

The scenarios that might occur in the linking process might be:

1). Shortened and nick names such as Cynthia and Cindy; Jennifer and Jenny; Margaret and Maggy; Patricia and Patty or Pat, were checked for possible matches, in all cases in Seattle, they turned out not to be the same person.

Because we turned up no cases where the same women used shortened and full name on two different official documents, we may conclude that a woman who uses her full first name on one official document would probably use it on all official documents. The same can be assumed for the shortened names; A woman who uses a shortened name on one official document, would probably use the same shortened name on all official documents.

A good portion of people may have changed addresses and names. The assumptions made for verifying addresses are:

2). A male with the same last name as the subject found at the same address (as reported in the CSS or birth records) as the subject is a spouse, or at least someone who would potentially move with the subject. In many cases, only the male is listed in Cole's Directories, making this assumption necessary for address checking.

3) The address given on the CSS records and the Birth Certificate records were considered to be the address where the subject would have resided in the given year. Therefore, the Cole's directory would have someone with the family name of the person at the given address.

4) The subject did not divorce if married in the CSS record and change names completely before giving birth to a child.

5) The subject actually gave birth to the child. (did not adopt). In Seattle, we may have the option of exploring this further and are in the process of looking into to this.